Canonical correlation analysis (II)

Reading:
AMSA: pages 550-566

Multivariate Analysis, Spring 2016
Institute of Statistics, National Chiao Tung University

May 31, 2016

Brief outline

1. Sample canonical variates and sample canonical correlations
2. Matrices of errors of approximations
3. Proportions of explained sample variance
4. Large sample inferences

Sample canonical variates and sample canonical correlations

- A random sample of n observations on each of the $(p+q)$ variables $\mathbf{X}^{(1)}$, $\mathbf{X}^{(2)}$ can be assembled into the $n \times (p+q)$ data matrix

$$
\begin{aligned}
\mathbf{X} &= \left[\mathbf{X}^{(1)} \,\middle|\, \mathbf{X}^{(2)} \right] \\
&= \left[\begin{array}{cccc|cccc}
x_{11}^{(1)} & x_{12}^{(1)} & \cdots & x_{1p}^{(1)} & x_{11}^{(2)} & x_{12}^{(2)} & \cdots & x_{1q}^{(2)} \\
x_{21}^{(1)} & x_{22}^{(1)} & \cdots & x_{2p}^{(1)} & x_{21}^{(2)} & x_{22}^{(2)} & \cdots & x_{2q}^{(2)} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
x_{n1}^{(1)} & x_{n2}^{(1)} & \cdots & x_{np}^{(1)} & x_{n1}^{(2)} & x_{n2}^{(2)} & \cdots & x_{nq}^{(2)}
\end{array}\right] \\
&= \left[\begin{array}{c|c}
\mathbf{x}_1^{(1)T} & \mathbf{x}_1^{(2)T} \\
\vdots & \vdots \\
\mathbf{x}_n^{(1)T} & \mathbf{x}_n^{(2)T}
\end{array}\right]
\end{aligned}
$$

- The vector of sample means

$$\bar{\mathbf{x}}_{(p+q)\times 1} = \left[\begin{array}{c} \bar{\mathbf{x}}^{(1)} \\ \hline \bar{\mathbf{x}}^{(2)} \end{array}\right]$$

where $\bar{\mathbf{x}}^{(1)} = \frac{1}{n}\sum_{j=1}^{n}\mathbf{x}_j^{(1)}$, $\bar{\mathbf{x}}^{(2)} = \frac{1}{n}\sum_{j=1}^{n}\mathbf{x}_j^{(2)}$

- The sample covariance matrix

$$\mathbf{S}_{(p+q)\times(p+q)} = \left[\begin{array}{c|c} \mathbf{S}_{11\ (p\times p)} & \mathbf{S}_{12\ (p\times q)} \\ \hline \mathbf{S}_{21\ (q\times p)} & \mathbf{S}_{22\ (q\times q)} \end{array}\right]$$

where $\mathbf{S}_{kl} = \frac{1}{n-1}\sum_{j=1}^{n}(\mathbf{x}_j^{(k)} - \bar{\mathbf{x}}^{(k)})(\mathbf{x}_j^{(l)} - \bar{\mathbf{x}}^{(l)})^T, k, l = 1, 2$

- The linear combinations

$$\widehat{U} = \widehat{\mathbf{a}}^T \mathbf{x}^{(1)} \qquad \widehat{V} = \widehat{\mathbf{b}}^T \mathbf{x}^{(2)}$$

have sample correlation

$$r_{\widehat{U},\widehat{V}} = \frac{s_{\widehat{U},\widehat{V}}}{\sqrt{s_{\widehat{U},\widehat{U}}}\sqrt{s_{\widehat{V},\widehat{V}}}} = \frac{\widehat{\mathbf{a}}^T \mathbf{S}_{12} \widehat{\mathbf{b}}}{\sqrt{\widehat{\mathbf{a}}^T \mathbf{S}_{11} \widehat{\mathbf{a}}}\sqrt{\widehat{\mathbf{b}}^T \mathbf{S}_{22} \widehat{\mathbf{b}}}} \tag{4}$$

- Define

$$
\begin{aligned}
\text{1st pair of sample} \quad &= \quad \text{the pair of linear combinations} \\
\text{canonical variates} \quad & \quad\;\; \widehat{U}_1 \text{ and } \widehat{V}_1 \text{ having unit sample variances,} \\
& \quad\;\; \text{which maximize the sample} \\
& \quad\;\; \text{correlation (4)} \\
\vdots \quad & \;\vdots \quad\;\; \vdots \\
k\text{th pair of sample} \quad &= \quad \text{the pair of linear combinations} \\
\text{canonical variates} \quad & \quad\;\; \widehat{U}_k \text{ and } \widehat{V}_k \text{ having unit sample variances,} \\
& \quad\;\; \text{which maximize the sample} \\
& \quad\;\; \text{correlation (4) among all choices} \\
& \quad\;\; \text{uncorrelated with the previous} \\
& \quad\;\; k-1 \text{ sample canonical variate pairs}
\end{aligned}
$$

- Let $\widehat{\rho_1^*}^2 \geq \widehat{\rho_2^*}^2 \geq \cdots \geq \widehat{\rho_p^*}^2$ be the eigenvalues of $\mathbf{S}_{11}^{-1/2}\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21}\mathbf{S}_{11}^{-1/2}$, and $\widehat{\mathbf{e}}_1, \widehat{\mathbf{e}}_2, \cdots, \widehat{\mathbf{e}}_p$ are the associated $(p \times 1)$ eigenvectors.

  $\widehat{\rho_1^*}^2, \widehat{\rho_2^*}^2, \cdots, \widehat{\rho_p^*}^2$ are also the p largest eigenvalues of $\mathbf{S}_{22}^{-1/2}\mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}\mathbf{S}_{22}^{-1/2}$ with corresponding $(q \times 1)$ eigenvectors $\widehat{\mathbf{f}}_1, \widehat{\mathbf{f}}_2, \cdots, \widehat{\mathbf{f}}_p$.

  $$\widehat{\mathbf{f}}_k = (1/\widehat{\rho_k^*})\mathbf{S}_{22}^{-1/2}\mathbf{S}_{21}\mathbf{S}_{11}^{-1/2}\widehat{\mathbf{e}}_k, k = 1, \cdots, p.$$

- Result: The kth sample canonical variate pair, $k = 1, 2, \cdots, p$,

$$\widehat{U}_k = \underbrace{\widehat{\mathbf{e}}_k^T \mathbf{S}_{11}^{-1/2}}_{\widehat{\mathbf{a}}_k^T} \mathbf{x}^{(1)} \text{ and } \widehat{V}_k = \underbrace{\widehat{\mathbf{f}}_k^T \mathbf{S}_{22}^{-1/2}}_{\widehat{\mathbf{b}}_k^T} \mathbf{x}^{(2)}$$

maximizes

$$r_{\widehat{U}_k, \widehat{V}_k} = \widehat{\rho_k^*}$$

among those linear combinations uncorrelated with the preceding $1, 2, \cdots, k-1$ sample canonical variate pairs.

- The sample canonical variates have the properties

$$\begin{aligned} s_{\widehat{U}_k,\widehat{U}_k} &= s_{\widehat{V}_k,\widehat{V}_k} = 1 \\ s_{\widehat{U}_k,\widehat{U}_\ell} &= s_{\widehat{V}_k,\widehat{V}_\ell} = 0 \quad k \neq \ell \\ s_{\widehat{U}_k,\widehat{V}_\ell} &= 0 \quad k \neq \ell \end{aligned}$$

for $k, \ell = 1, 2, \cdots, p$.

- The interpretation of $\widehat{U}_k$, $\widehat{V}_k$ is often aided by computing the sample correlations between the canonical variates and the variables in the sets $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$.
- Let $\widehat{\mathbf{A}}_{(p\times p)} = [\widehat{\mathbf{a}}_1, \cdots, \widehat{\mathbf{a}}_p]^T$ and $\widehat{\mathbf{B}}_{(q\times q)} = [\widehat{\mathbf{b}}_1, \cdots, \widehat{\mathbf{b}}_q]^T$, so that the vector of sample canonical variables are

$$\widehat{\mathbf{U}}_{(p\times 1)} = \widehat{\mathbf{A}}\mathbf{x}^{(1)} \quad \widehat{\mathbf{V}}_{(q\times 1)} = \widehat{\mathbf{B}}\mathbf{x}^{(2)},$$

  where we are primarily interested in the first p canonical variables in $\widehat{\mathbf{V}}$.

- Then, we have

$$
\begin{aligned}
\mathbf{R}_{\widehat{\mathbf{U}},\mathbf{x}^{(1)}} &= \text{matrix of sample correlations of } \widehat{\mathbf{U}} \text{ with } \mathbf{x}^{(1)} = \widehat{\mathbf{A}}\mathbf{S}_{11}\mathbf{D}_{11}^{-1/2} \\
\mathbf{R}_{\widehat{\mathbf{V}},\mathbf{x}^{(2)}} &= \text{matrix of sample correlations of } \widehat{\mathbf{V}} \text{ with } \mathbf{x}^{(2)} = \widehat{\mathbf{B}}\mathbf{S}_{22}\mathbf{D}_{22}^{-1/2} \\
\mathbf{R}_{\widehat{\mathbf{U}},\mathbf{x}^{(2)}} &= \text{matrix of sample correlations of } \widehat{\mathbf{U}} \text{ with } \mathbf{x}^{(2)} = \widehat{\mathbf{A}}\mathbf{S}_{12}\mathbf{D}_{22}^{-1/2} \\
\mathbf{R}_{\widehat{\mathbf{V}},\mathbf{x}^{(1)}} &= \text{matrix of sample correlations of } \widehat{\mathbf{V}} \text{ with } \mathbf{x}^{(1)} = \widehat{\mathbf{B}}\mathbf{S}_{21}\mathbf{D}_{11}^{-1/2}
\end{aligned}
$$

where $\mathbf{D}_{11} = \text{diag}(\mathbf{S}_{11})$ and $\mathbf{D}_{22} = \text{diag}(\mathbf{S}_{22})$.

Sample canonical variates from standardized variables

- If the observations are standardized, the data matrix becomes

$$
\begin{aligned}
\mathbf{Z} &= \left[\begin{array}{c|c} \mathbf{Z}^{(1)} & \mathbf{Z}^{(2)} \end{array} \right] \\
&= \left[\begin{array}{c|c} \mathbf{z}_1^{(1)T} & \mathbf{z}_1^{(2)T} \\ \vdots & \vdots \\ \mathbf{z}_n^{(1)T} & \mathbf{z}_n^{(2)T} \end{array} \right] \\
&= \left[\begin{array}{c|c} (\mathbf{x}_1^{(1)T} - \overline{\mathbf{x}}^{(1)T})\mathbf{D}_{11}^{-1/2} & (\mathbf{x}_1^{(2)T} - \overline{\mathbf{x}}^{(2)T})\mathbf{D}_{22}^{-1/2} \\ \vdots & \vdots \\ (\mathbf{x}_n^{(1)T} - \overline{\mathbf{x}}^{(1)T})\mathbf{D}_{11}^{-1/2} & (\mathbf{x}_n^{(2)T} - \overline{\mathbf{x}}^{(2)T})\mathbf{D}_{22}^{-1/2} \end{array} \right]
\end{aligned}
$$

- The sample canonical variates from standardized variables can be derived from the **sample correlation matrices** in a manner consistent with the population case described previously.

Sample descriptive measures of goodness-of-fit

- It is useful to have summary measures of the extent to which the canonical variates account for the variation in their respective sets.
  - If the canonical variates are good summaries of their respective sets of variables, the the association between variables can be described in terms of the canonical variates and their correlations.
- Two sample descriptive measures:
  1. matrices of errors of approximations
  2. proportions of explained sample variance

Matrices of errors of approximations

- Since

$$\mathbf{x}^{(1)}_{(p\times 1)} = \widehat{\mathbf{A}}^{-1}_{(p\times p)}\widehat{\mathbf{U}}_{(p\times 1)} \quad \mathbf{x}^{(2)}_{(q\times 1)} = \widehat{\mathbf{B}}^{-1}_{(q\times q)}\widehat{\mathbf{V}}_{(q\times 1)}$$

we have

$$\begin{aligned}
\mathbf{S}_{12} &= \widehat{\mathbf{A}}^{-1}\left[\begin{array}{cccc|c} \widehat{\rho_1^*} & 0 & \cdots & 0 & \\ 0 & \widehat{\rho_2^*} & \cdots & 0 & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \\ 0 & 0 & \cdots & \widehat{\rho_p^*} & \end{array}\right](\widehat{\mathbf{B}}^{-1})^T \\
&= \widehat{\rho_1^*}\widehat{\mathbf{a}}^{(1)}\widehat{\mathbf{b}}^{(1)T} + \widehat{\rho_2^*}\widehat{\mathbf{a}}^{(2)}\widehat{\mathbf{b}}^{(2)T} + \cdots + \widehat{\rho_p^*}\widehat{\mathbf{a}}^{(p)}\widehat{\mathbf{b}}^{(p)T} \\
\mathbf{S}_{11} &= (\widehat{\mathbf{A}}^{-1})(\widehat{\mathbf{A}}^{-1})^T = \widehat{\mathbf{a}}^{(1)}\widehat{\mathbf{a}}^{(1)T} + \widehat{\mathbf{a}}^{(2)}\widehat{\mathbf{a}}^{(2)T} + \cdots + \widehat{\mathbf{a}}^{(p)}\widehat{\mathbf{a}}^{(p)T} \\
\mathbf{S}_{22} &= (\widehat{\mathbf{B}}^{-1})(\widehat{\mathbf{B}}^{-1})^T = \widehat{\mathbf{b}}^{(1)}\widehat{\mathbf{b}}^{(1)T} + \widehat{\mathbf{b}}^{(2)}\widehat{\mathbf{b}}^{(2)T} + \cdots + \widehat{\mathbf{b}}^{(q)}\widehat{\mathbf{b}}^{(q)T}
\end{aligned}$$

where $\widehat{\mathbf{a}}^{(i)}$ and $\widehat{\mathbf{b}}^{(i)}$ denote the ith column of $\widehat{\mathbf{A}}^{-1}$ and $\widehat{\mathbf{B}}^{-1}$, respectively.

- If only the first r canonical pairs are used, we have

$$\widetilde{\mathbf{x}}^{(1)}_{(p\times 1)} = \left[\ \widehat{\mathbf{a}}^{(1)} \mid \widehat{\mathbf{a}}^{(2)} \mid \cdots \mid \widehat{\mathbf{a}}^{(r)}\ \right] \begin{bmatrix} \widehat{U}_1 \\ \widehat{U}_2 \\ \vdots \\ \widehat{U}_r \end{bmatrix}$$

$$\widetilde{\mathbf{x}}^{(2)}_{(q\times 1)} = \left[\ \widehat{\mathbf{b}}^{(1)} \mid \widehat{\mathbf{b}}^{(2)} \mid \cdots \mid \widehat{\mathbf{b}}^{(r)}\ \right] \begin{bmatrix} \widehat{V}_1 \\ \widehat{V}_2 \\ \vdots \\ \widehat{V}_r \end{bmatrix}$$

then
$\mathbf{S}_{12}$ is approximate by sample $\mathrm{Cov}(\widetilde{\mathbf{x}}^{(1)}, \widetilde{\mathbf{x}}^{(2)})$,
$\mathbf{S}_{11}$ is approximate by sample $\mathrm{Cov}(\widetilde{\mathbf{x}}^{(1)})$,
$\mathbf{S}_{22}$ is approximate by sample $\mathrm{Cov}(\widetilde{\mathbf{x}}^{(2)})$.

- We see that the **matrices of errors of approximations**

$$\mathbf{S}_{11} - (\widehat{\mathbf{a}}^{(1)}\widehat{\mathbf{a}}^{(1)T} + \cdots + \widehat{\mathbf{a}}^{(r)}\widehat{\mathbf{a}}^{(r)T}) = \widehat{\mathbf{a}}^{(r+1)}\widehat{\mathbf{a}}^{(r+1)T} + \cdots + \widehat{\mathbf{a}}^{(p)}\widehat{\mathbf{a}}^{(p)T}$$

$$\mathbf{S}_{22} - (\widehat{\mathbf{b}}^{(1)}\widehat{\mathbf{b}}^{(1)T} + \cdots + \widehat{\mathbf{b}}^{(r)}\widehat{\mathbf{b}}^{(r)T}) = \widehat{\mathbf{b}}^{(r+1)}\widehat{\mathbf{b}}^{(r+1)T} + \cdots + \widehat{\mathbf{b}}^{(q)}\widehat{\mathbf{b}}^{(q)T}$$

$$\mathbf{S}_{12} - (\widehat{\rho_1^*}\widehat{\mathbf{a}}^{(1)}\widehat{\mathbf{b}}^{(1)T} + \cdots + \widehat{\rho_r^*}\widehat{\mathbf{a}}^{(r)}\widehat{\mathbf{b}}^{(r)T}) = \widehat{\rho_{r+1}^*}\widehat{\mathbf{a}}^{(r+1)}\widehat{\mathbf{b}}^{(r+1)T} + \cdots + \widehat{\rho_p^*}\widehat{\mathbf{a}}^{(p)}\widehat{\mathbf{b}}^{(p)T}$$

- The matrices of errors of approximations may be interpreted as descriptive summaries of how well the first r sample canonical variates reproduce the sample covariance matrices.
- Large entries in the rows and/or columns of the approximation error matrices indicate a **poor fit** to the corresponding variable(s).

Proportions of explained sample variance

- When the observations are standardized, the sample covariance matrices $\mathbf{S}_{kl}$ are the sample correlation matrices $\mathbf{R}_{kl}$, and let the canonical coefficient matrices be $\widehat{\mathbf{A}}_{\mathbf{z}}$ and $\widehat{\mathbf{B}}_{\mathbf{z}}$.

- Since

$$
\begin{aligned}
\text{sample Corr}(\mathbf{z}^{(1)}, \widehat{\mathbf{U}}) &= \text{sample Cov}(\mathbf{z}^{(1)}, \widehat{\mathbf{U}}) \\
&= \text{sample Cov}(\widehat{\mathbf{A}}_{\mathbf{z}}^{-1}\widehat{\mathbf{U}}, \widehat{\mathbf{U}}) = \widehat{\mathbf{A}}_{\mathbf{z}}^{-1} \\
\text{sample Corr}(\mathbf{z}^{(2)}, \widehat{\mathbf{V}}) &= \text{sample Cov}(\mathbf{z}^{(2)}, \widehat{\mathbf{V}}) \\
&= \text{sample Cov}(\widehat{\mathbf{B}}_{\mathbf{z}}^{-1}\widehat{\mathbf{V}}, \widehat{\mathbf{V}}) = \widehat{\mathbf{B}}_{\mathbf{z}}^{-1}
\end{aligned}
$$

So

$$
\widehat{\mathbf{A}}_{\mathbf{z}}^{-1} = [\widehat{\mathbf{a}}_{\mathbf{z}}^{(1)}, \widehat{\mathbf{a}}_{\mathbf{z}}^{(2)}, \cdots, \widehat{\mathbf{a}}_{\mathbf{z}}^{(p)}] = \begin{bmatrix} r_{\widehat{U}_1, z_1^{(1)}} & r_{\widehat{U}_2, z_1^{(1)}} & \cdots & r_{\widehat{U}_p, z_1^{(1)}} \\ r_{\widehat{U}_1, z_2^{(1)}} & r_{\widehat{U}_2, z_2^{(1)}} & \cdots & r_{\widehat{U}_p, z_2^{(1)}} \\ \vdots & \vdots & \vdots & \vdots \\ r_{\widehat{U}_1, z_p^{(1)}} & r_{\widehat{U}_2, z_p^{(1)}} & \cdots & r_{\widehat{U}_p, z_p^{(1)}} \end{bmatrix}
$$

$$
\widehat{\mathbf{B}}_{\mathbf{z}}^{-1} = [\widehat{\mathbf{b}}_{\mathbf{z}}^{(1)}, \widehat{\mathbf{b}}_{\mathbf{z}}^{(2)}, \cdots, \widehat{\mathbf{b}}_{\mathbf{z}}^{(q)}] = \begin{bmatrix} r_{\widehat{V}_1, z_1^{(2)}} & r_{\widehat{V}_2, z_1^{(2)}} & \cdots & r_{\widehat{V}_q, z_1^{(2)}} \\ r_{\widehat{V}_1, z_2^{(2)}} & r_{\widehat{V}_2, z_2^{(2)}} & \cdots & r_{\widehat{V}_q, z_2^{(2)}} \\ \vdots & \vdots & \vdots & \vdots \\ r_{\widehat{V}_1, z_q^{(2)}} & r_{\widehat{V}_2, z_q^{(2)}} & \cdots & r_{\widehat{V}_q, z_q^{(2)}} \end{bmatrix}
$$

- Total (standardized) sample variance in the first set

$$= \mathrm{tr}(\mathbf{R}_{11}) = \mathrm{tr}(\widehat{\mathbf{a}}_{\mathbf{z}}^{(1)}\widehat{\mathbf{a}}_{\mathbf{z}}^{(1)T} + \widehat{\mathbf{a}}_{\mathbf{z}}^{(2)}\widehat{\mathbf{a}}_{\mathbf{z}}^{(2)T} + \cdots + \widehat{\mathbf{a}}_{\mathbf{z}}^{(p)}\widehat{\mathbf{a}}_{\mathbf{z}}^{(p)T}) = p$$

- Total (standardized) sample variance in the second set

$$= \mathrm{tr}(\mathbf{R}_{22}) = \mathrm{tr}(\widehat{\mathbf{b}}_{\mathbf{z}}^{(1)}\widehat{\mathbf{b}}_{\mathbf{z}}^{(1)T} + \widehat{\mathbf{b}}_{\mathbf{z}}^{(2)}\widehat{\mathbf{b}}_{\mathbf{z}}^{(2)T} + \cdots + \widehat{\mathbf{b}}_{\mathbf{z}}^{(q)}\widehat{\mathbf{b}}_{\mathbf{z}}^{(q)T}) = q$$

- If only the first r canonical pairs are used, we define the contributions of the first r canonical variates to the total (standardized) sample variances as

$$\mathrm{tr}(\widehat{\mathbf{a}}_{\mathbf{z}}^{(1)}\widehat{\mathbf{a}}_{\mathbf{z}}^{(1)T} + \widehat{\mathbf{a}}_{\mathbf{z}}^{(2)}\widehat{\mathbf{a}}_{\mathbf{z}}^{(2)T} + \cdots + \widehat{\mathbf{a}}_{\mathbf{z}}^{(r)}\widehat{\mathbf{a}}_{\mathbf{z}}^{(r)T}) = \sum_{i=1}^{r}\sum_{k=1}^{p} r^2_{\widehat{U}_i, z_k^{(1)}}$$

$$\mathrm{tr}(\widehat{\mathbf{b}}_{\mathbf{z}}^{(1)}\widehat{\mathbf{b}}_{\mathbf{z}}^{(1)T} + \widehat{\mathbf{b}}_{\mathbf{z}}^{(2)}\widehat{\mathbf{b}}_{\mathbf{z}}^{(2)T} + \cdots + \widehat{\mathbf{b}}_{\mathbf{z}}^{(r)}\widehat{\mathbf{b}}_{\mathbf{z}}^{(r)T}) = \sum_{i=1}^{r}\sum_{k=1}^{q} r^2_{\widehat{V}_i, z_k^{(2)}}$$

- The proportions of total (standardized) sample variances **explained** by the first r canonical variates become

$$
\begin{aligned}
R_{\mathbf{z}^{(1)}|\widehat{U}_1,\cdots,\widehat{U}_r} &= \begin{pmatrix} \text{proportion of total standardized} \\ \text{sample variance in first set} \\ \text{explained by } \widehat{U}_1,\cdots,\widehat{U}_r \end{pmatrix} \\
&= \frac{\mathrm{tr}(\widehat{\mathbf{a}}_{\mathbf{z}}^{(1)}\widehat{\mathbf{a}}_{\mathbf{z}}^{(1)T}+\cdots+\widehat{\mathbf{a}}_{\mathbf{z}}^{(r)}\widehat{\mathbf{a}}_{\mathbf{z}}^{(r)T})}{\mathrm{tr}(\mathbf{R}_{11})} \\
&= \frac{\sum_{i=1}^{r}\sum_{k=1}^{p} r^2_{\widehat{U}_i,z_k^{(1)}}}{p} \\
R_{\mathbf{z}^{(2)}|\widehat{V}_1,\cdots,\widehat{V}_r} &= \begin{pmatrix} \text{proportion of total standardized} \\ \text{sample variance in second set} \\ \text{explained by } \widehat{V}_1,\cdots,\widehat{V}_r \end{pmatrix} \\
&= \frac{\mathrm{tr}(\widehat{\mathbf{b}}_{\mathbf{z}}^{(1)}\widehat{\mathbf{b}}_{\mathbf{z}}^{(1)T}+\cdots+\widehat{\mathbf{b}}_{\mathbf{z}}^{(r)}\widehat{\mathbf{b}}_{\mathbf{z}}^{(r)T})}{\mathrm{tr}(\mathbf{R}_{22})} \\
&= \frac{\sum_{i=1}^{r}\sum_{k=1}^{q} r^2_{\widehat{V}_i,z_k^{(2)}}}{q}
\end{aligned}
$$

Large sample inferences

- The likelihood ration test for

$$H_0 : \mathbf{\Sigma}_{12} = \mathbf{0} \quad (\rho_1^* = \rho_2^* = \cdots = \rho_p^* = 0)$$

Reject H_0 at significant level α if

$$-2 \ln \Lambda = n \ln \left(\frac{|\mathbf{S}_{11}||\mathbf{S}_{22}|}{|\mathbf{S}|} \right) = -n \ln \prod_{i=1}^{p} (1 - \widehat{\rho_i^*}^2) > \chi^2_{pq}(\alpha),$$

where $\chi^2_{pq}(\alpha)$ is the upper (100α)th percentile of a chi-square distribution with d.f.$=pq$.

  - Bartlett suggests replacing the multiplicative factor n with the factor $n - 1 - \frac{1}{2}(p + q + 1)$ to improve the χ^2 approximation.

- To test

$$
\begin{aligned}
H_0^{(k)} &: \rho_1^* \neq 0, \rho_2^* \neq 0, \cdots, \rho_k^* \neq 0, \rho_{k+1}^* = \cdots = \rho_p^* = 0 \\
H_1^{(k)} &: \rho_i^* \neq 0, \text{ for some } i \geq k+1
\end{aligned}
$$

Reject $H_0^{(k)}$ at significant level α if

$$
-(n-1-\frac{1}{2}(p+q+1)) \ln \prod_{i=k+1}^{p} (1-\widehat{\rho_i^*}^2) > \chi^2_{(p-k)(q-k)}(\alpha)
$$

- If $H_0 : \boldsymbol{\Sigma}_{12} = \mathbf{0}$ $(\rho_1^* = \rho_2^* = \cdots = \rho_p^* = 0)$ is rejected, it is natural to examine the **significance** of the individual canonical correlations. We can test the sequence of hypotheses:
    1. Begin from testing $H_0^{(1)}$ (assuming that the 1st canonical correlation is nonzero and the remaining $p-1$ canonical correlations are zero)
    2. If $H_0^{(1)}$ is rejected, then test $H_0^{(2)}$ (assuming that the first two canonical correlation is nonzero and the remaining $p-2$ canonical correlations are zero)
    3. And so forth.